1. What is TensorFlow? Which company is the leading contributor to TensorFlow?

Tensorflow is an open-source library used primarily for machine learning and AI development. As its name implies, the library provides many tools to work with tensors, the basis of neural networks and deep learning. Google is the leading contributor to TensorFlow

2. What is TensorRT? How is it different from TensorFlow?

TensorRT is an SDK by NVIDIA that works with in conjunction with Tensorflow. It is a model optimizer that can reduce training time while improving accuracy and performance in production for models that are being developed on Tensorflow. One of its key abilities is to fully leverage the power of GPUs that the algorithm might be running on

3. What is ImageNet? How many images does it contain? How many classes?

ImageNet is an image database based on WordNet. In WordNet, a group of words (mostly nouns) are grouped by how well they can describe a concept. These groupings are called synsets. ImageNet provides about 1000 images per synset. There are over 100 000 synsets therefore likely 100 000 000+ images. As of April 30th, 2010, the statistics are:

* Total number of non-empty synsets: 21841
* Total number of images: 14,197,122
* Number of images with bounding box annotations: 1,034,908
* Number of synsets with SIFT features: 1000
* Number of images with SIFT features: 1.2 million

Additionally, there are 27 categories including sports, birds, animals, etc.

4. Please research and explain the differences between MobileNet and GoogleNet (Inception) architectures.

MobileNets are a family of models built on Tensorflow and optimized for mobile devices. There are two hyperparameters that you can tune that allow you to adjust between latency and accuracy. GoogLeNet is a pretrained convolutional neural network that is 22 layers deep and is something you can just plug data in. The network trained on ImageNet classifies images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

5. In your own words, what is a bottleneck?

A bottleneck is a restricting gateway or process that constrains the volume or velocity of something as it passes through.

6. How is a bottleneck different from the concept of layer freezing?

From a machine learning perspective, bottlenecking is when you pass data through a pretrained network/weights that are frozen. The information is then passed through an unfrozen network where weights can be adjusted. Layer freezing is essentially where the weights of a layer are unchanged after the model is trained in prior iterations.

7. In this lab, you trained the last layer (all the previous layers retain their already-trained state). Explain how the lab used the previous layers (where did they come from? how were they used in the process?)

8. How does a low --learning\_rate (step 7) value (like 0.005) affect the precision? How much longer does training take?

9. How about a --learning\_rate (step 7) of 1.0? Is the precision still good enough to produce a usable graph?

10. For step 8, you can use any images you like. Pictures of food, people, or animals work well. You can even use ImageNet images. How accurate was your model? Were you able to train it using a few images, or did you need a lot?

11. Run the script on the CPU (see instructions above) How does the training time compare to the default network training (section 4)? Why?

12. Try the training again, but this time do export ARCHITECTURE="inception\_v3" Are CPU and GPU training times different?

13. Given the hints under the notes section, if we trained Inception\_v3, what do we need to pass to replace ??? below to the label\_image script? Can we also glean the answer from examining TensorBoard?